Issues in Information Systems
Volume 23, Issue 4, pp. 254-264, 2022

DOI: https://doi.org/10.48009/4_iis_2022_122

Teaching Data Warehousing with SAP HANA

Ming Wang, *California State University, mwang3@calstatela.edu*

Abstract

SAP HANA is an innovative in-memory enterprise data platform with advanced data warehouse characteristics. Many DBMS vendors have recently updated their products with data-in-memory and columnar data storage features in their new versions. To catch the database's industrial trend, the author has introduced SAP HANA technology into the database and data warehousing courses using SAP University Alliances curriculum materials. The author has taught SAP HANA for data warehousing in her Database and Data Warehouse Systems course for over three years. SAP HANA course materials are downloaded from SAP University Alliances Learning Portal because the Data Warehouse Chapter in the textbook does not cover specific SAP HANA features. The paper presents the importance, framework, and delivery models of integrating SAP HANA technology into this course. The study focuses on the data warehouse architecture, components, data storage, modeling, and implementation. SAP HANA Business Data Warehouse (BW), business intelligence (BI) course materials. Student learning outcomes of SAP HANA are assessed and analyzed. Learning by implementing data housing systems is vital in teaching the next generation of digital business leaders. The student engagement with hands-on activities needs to be on the real-world enterprise data platform.

Keywords: IS Curriculum, Database, Data Warehouse, SAP HANA

Introduction

The databases and data warehouses are two primary data sources for business intelligence and analytics. Traditionally, they are defined as two different types of software products and serve different data processing needs within organizations. Databases are developed for online transactional processing (OLTP), while data warehouses are designed for online analytical processing (OLAP). But the recent development of SAP HANA technology has combined these two primary data sources into one big data in-memory platform. The In-Memory Column Store in SAP HANA has made the trend in the database and data warehouse industries. Many database vendors have recently updated their new versions of database management systems (DBMSs) with data-in-memory and columnar data storage features. For example, Oracle 12 has made evolutional changes by allowing data to be stored either in rows or in columns and running them in memory.

To catch up with the industrial trend and meet the challenges, the author has taught the Database and Data Warehouse Systems course at undergraduate and graduate levels for over three years. She uses Oracle DBMS to teach relational/object-relational databases and SAP HANA to teach data warehousing. It is important and appropriate for students to learn databases before data warehouses and warehousing based on the author's experience. In this paper, the author proposes the teaching SAP HANA framework, demonstrates how to teach the specific modules in this framework, and finally provides the assessment of student learning outcomes of the framework.

Data warehousing is an important technology that has been adopted and utilized by more and more organizations. Teaching data warehousing in IS curriculum with hands-on experience with the software in the real world will get students better prepared to be next-generation digital business leaders. The paper aims to provide IS educators with an alternative way to teach advanced data warehousing technology.

Framework

The author proposes the teaching SAP HANA framework in Figure 1 based on her teaching experience in the area. It takes one week to teach each of the topics in the framework. The assumption is that students already have learned the database knowledge and skills as prerequisites.

Module	**Content**
Module 1	Introducing Data Warehousing and Data Integration Concepts
Module 2	Comparing data warehouse architectures using SAP BW and SAP HANA
Module 3	Implementing SAP HANA on ERPsim Games and SQL Query on the results
Module 4	Multi-dimensional data modeling and analytics on HANA in ERPsim via OData service
Module 5	Developing executive analytics project on data generated from ERPsim Games

Figure 1 Teaching Data Warehousing with SAP HANA Framework

Student Learning Objectives are as follows.

- Distinguish databases and data warehouses
- Distinguish transactional systems and informational systems
- Distinguish OLTP and OLAP
- Evaluate and compare three Data Warehouse Architectures
- Learn dependent data warehouse architecture implementing SAP BW
- Learn Logical Data Mart and Real-Time Data Warehouse Architecture by implementing SAP HANA
- Model data using star schema and snowflake schema
- Implement Structured Query language (SQL) queries on HANA via OData in the ERPsim system
- Perform Multi-dimensional data modeling and analytics on HANA in the ERPsim system

Teaching Modules

The author proposes teaching data warehousing with SAP HANA modules in this section.

Module 1. Introducing Data Warehousing and Data Integration Concepts

1. The architecture for data warehouses has evolved into three different variations: the independent data mart architecture, dependent data mart & operational data store architecture, and logical data mart and real-time data warehouse architecture. The comparison of the Three Data Warehouse Architectures is as follows.

Independent Data Mart Architecture is supported by Kimball (1997). Figure 2 (Hoffer, Ramesh, & Topi, 2022) shows the independent data mart architecture of data warehouses. The data warehouse is the collection of data marts. Data are extracted from the multiple internal and external source system files and databases into numerous data marts. The data from the various source systems are transformed and integrated before being loaded into the data marts. Data marts are mini warehouses, limited in scope. Separate ETL (Exact, Transform and Load) is for each independent data mart. Data access complexity of the data warehouse is due to multiple data marts. The independent data mart architecture in Figure 2 has several significant limitations (Marco, 2000; Meyer, 1997) and Inmon (1997, 2000) points out the fallacies.

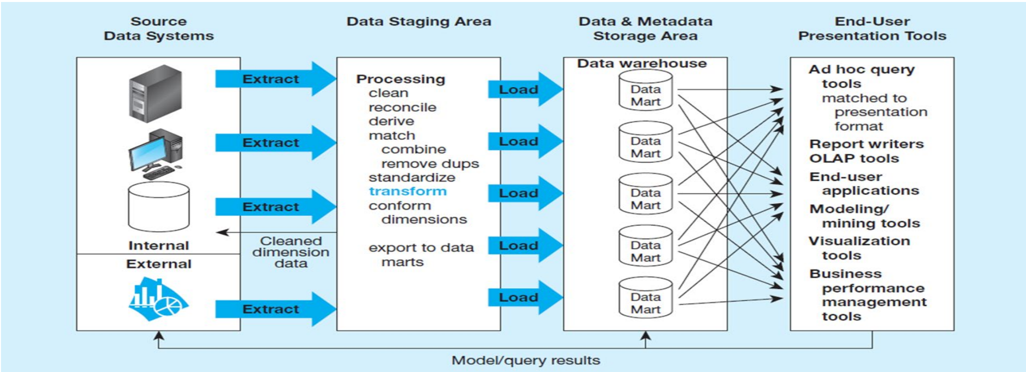

Figure 2 Independent Data Mart Architecture, Hoffer, Ramesh, & Topi, (2022)

- The separate ETL process for each data mart leads to redundant data and processing efforts
- Inconsistency between data marts
- Challenging to drill down for related facts between data marts
- Excessive scaling costs the more applications to be developed
- High cost for obtaining consistency between marts

Dependent Data Mart & Operational Data Store (ODS) Architecture uses a three-level approach represented by the dependent data mart and operational data store architecture in Figure 3 (Hoffer, Ramesh, & Topi, 2022).

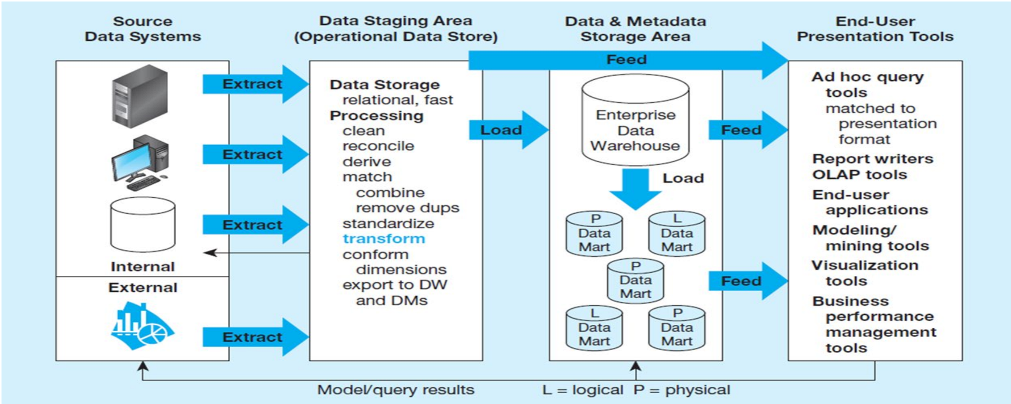

Figure 3 Dependent Data Mart and Operational Data Store Hoffer, Ramesh, & Topi, (2022)

The new level is the operational-data store. Operational-data store provides an option for transforming current data from different source data systems. A single ETL is used for enterprise data warehouse (EDW). The dependent data marts are loaded from an enterprise data warehouse (EDW), a central, integrated data warehouse that is the control point, and a single "version of the truth" made available to end users for decision support applications. Bill Inmon (1997), a data warehousing pioneer, supports this more centralized approach. Dependent data marts serve the decision-making needs of business processes for user groups. A user group can access either its data mart or other data in EDW. Integration of data is the responsibility of the enterprise data warehouse architects; An operational data store (ODS) is an integrated, subject-oriented, continuously updateable, current-valued (with recent history), and organization-wide, detailed database designed to serve operational users.

Logical Data Mart & Real-Time Data Warehouse Architecture **(Hoffer, Ramesh, & Topi, 2022)** means that the source data systems, decision support services, and the data warehouse exchange data perform near-real-time because of big data in-memory technology. Logical data marts are not physically separate databases, but different relational views of one physical slightly denormalized relational data warehouse. Data are moved into the data warehouse with the high-performance computing power of the warehouse technology. New data marts can be created quickly because no physical database or database technology needs to be developed or acquired, and no loading routines need to be written. A relational view of a data warehouse creates a logical data mart. Figure 4 illustrates the logical data mart and real-time data warehouse architecture

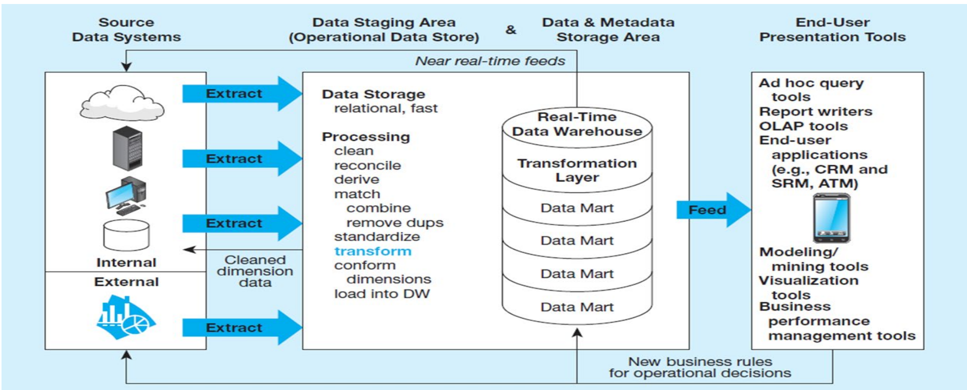

Figure 4 Logical Data Mart and Real-Time Data Warehouse, Hoffer, Ramesh, & Topi, (2022)

Module 2. Comparing data warehouse architectures using SAP BW and SAP HANA

Traditional SAP Business Data warehouse (BW) uses the dependent data mart and operational data store architecture. Students conduct the following activities

1. Identify source system(s)
2. Connect source system
3. Build data marts/info cubes
   a) Build and activate SAP BI InfoObjects, InfoProviders, etc.
   b) Build and activate extractors to create extraction structures (DataSource in source system)
   c) Replicate DataSources into SAP BI
   d) Migrate, as needed, from 3. x to 7.0 data flow
4. Perform
   a) Initial ETL for master data
   b) Initial ETL for transactional data
   c) Master data maintenance
   d) Transactional data loads

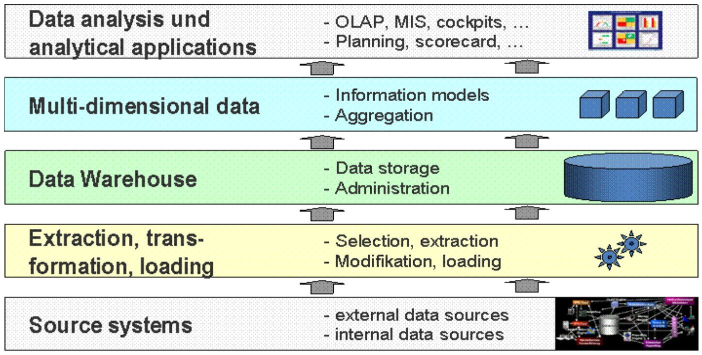

Figure 5. SAP BW architecture, Copyright SAP AG, Klaus Freyburger and Peter Lehmann

Figure 6 indicates that access to information can be either from the operational data store or dependent data marts in SAP BW.

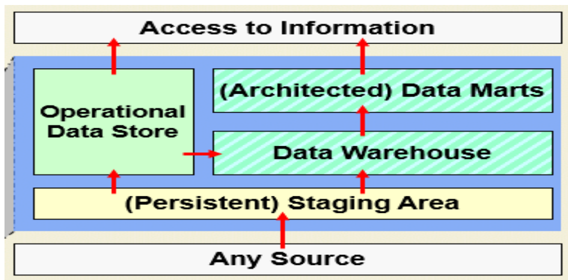

Figure 6 Access to information in SAP BW, Copyright SAP AG, Klaus Freyburger and Peter Lehmann

SAP HANA is an innovative data in-memory platform system using the Logical Data Mart & Real-Time Data Warehouse architecture. Figure 7 illustrates the architecture overview of SAP HANA.

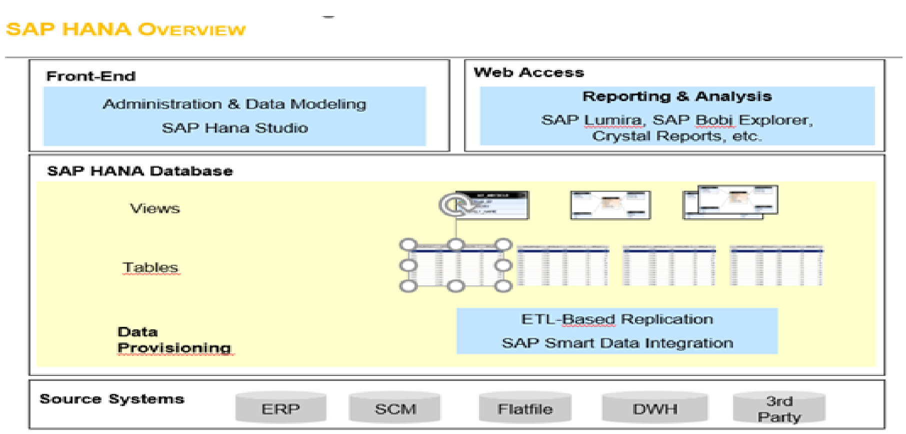

Figure 7 Architecture Overview of SAP HANA, Copyright SAP AG, Klaus Freyburger and Peter Lehmann

SAP HANA stores data in the columnar table for faster data retrieval speed. See the illustration in Figure 8. The four records in the Row-store table can be stored in the columnar table as follows.

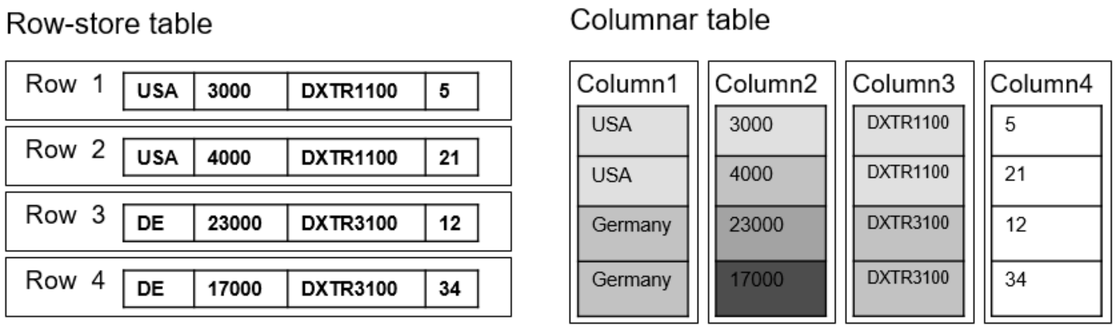

Figure 8 Data Storage in SAP HANA

SAP HANA ‘s underlying columnar in-memory relational database is exceptionally fast because of the index on the columns. In a columnar database, all values in a specific column are stored consecutively at one location, resulting in significantly faster access than in the traditional row-store-oriented database. This storage and access mechanism enables much faster aggregation operations like Sum, Min, Max, Count, and Average, Varma, N. (2018). SAP HANA in-memory appliance allows larger data volumes to be parallelly processed in real-time and optimizes analytics. It supports multi-tenants and runs applications either on the cloud or on-premises. Smart data Access can integrate tables from third-party database systems like Oracle and MS SQL Server. IBM DB2and Hadoop via Smart Data Access (Matthias, M.,Hügens, T. , Blum, S. (2017). Many applications and tools, such as standard SQL and MDX, can be integrated with SAP HANA.

SAP HANA analytic applications use three views: Attribute Views, Analytic Views, and Calculation Views. Attribute views represent master data (attributes, texts, hierarchies) and provide reusable dimensions for analytic and calculation views; Analytic Views join facts with relevant attribute dimensions. Calculation Views address more complex requirements than Analytic Views and can include both tables and views.

Module 3. Implementing SAP HANA on ERPsim Games and SQL Query on the results

In the phase, students are divided into teams to play ERPsim games. Each team member is assigned a role for each of the business processes. Three business roles in the ERPsim Distribution Game are illustrated in Figure 9: 1) CEO - Team Leader in charge of the company watching Financial Accounting, 2) Material Manager in charge of material planning and purchase, and 3) Sales and Distribution Manager in charge of marketing and price justification.

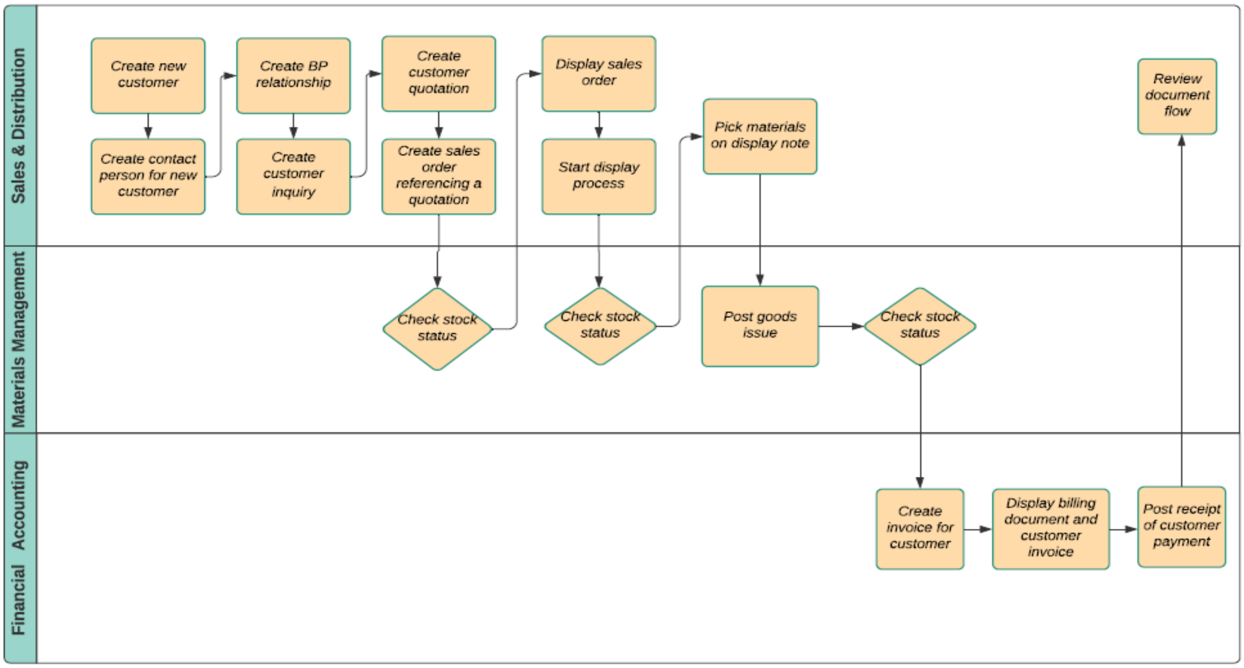

Figure 9. Student Roles in Each Team in ERPsim Games

Each team runs a company with simulated interactions between suppliers and customers in the cash-to-cash cycle in the beverage distribution enterprise. In the gamified environment, student players can visualize and supervise data changes in real-time on SAP HANA using SAP Predictive Analytics and make decisions on their business strategies. The winning team will have the highest company valuation and revenue.

Students can quickly learn SQL in SAP HANA based on SQL in relational databases. The following SQL statement shows SQL on HANA for ERP Distribution Game using OData 2.0 Service. The differences are the entity name has the data link prefix, and column names are case sensitive in the SQL syntax in HANA.

```
--1. List Customer numbers, cities, regions, and their corresponding cost and net revenue.
select CUSTOMER_NUMBER, CITY, REGION, COST, NET_VALUE
from "ERPsim.OData.erpsim.erpsim"."entities"."Sales";
```

Module 4. Multi-dimensional data modeling and analytics on HANA in ERPsim via OData service

The multi-dimensional modeling capacities were built directly into SAP HANA when data prevision was designed and developed. Students can model and analyze data generated from ERPsim Distribution Game (Léger et al., 2017 and Michon, et al., 2017).

The star schema is a data modeling diagram for data warehouse design. Dimensional (or descriptive) data are separated from fact or event data in fact tables. The Sales star schema in Figure 10 consists of one fact table containing event data and links to dimensions. Dimension tables contain descriptive information. For example, we might have a dimension table for the Sales entity with a unique ID and several granularity levels. There might be an entity description, state, city, zip code, etc. The dimension keys are also in the fact table. Therefore, we can do drill-down type queries. Figure 10 illustrates the Sales star schema for Multi-dimensional data modeling of HANA Data Warehouse in ERPsim games.

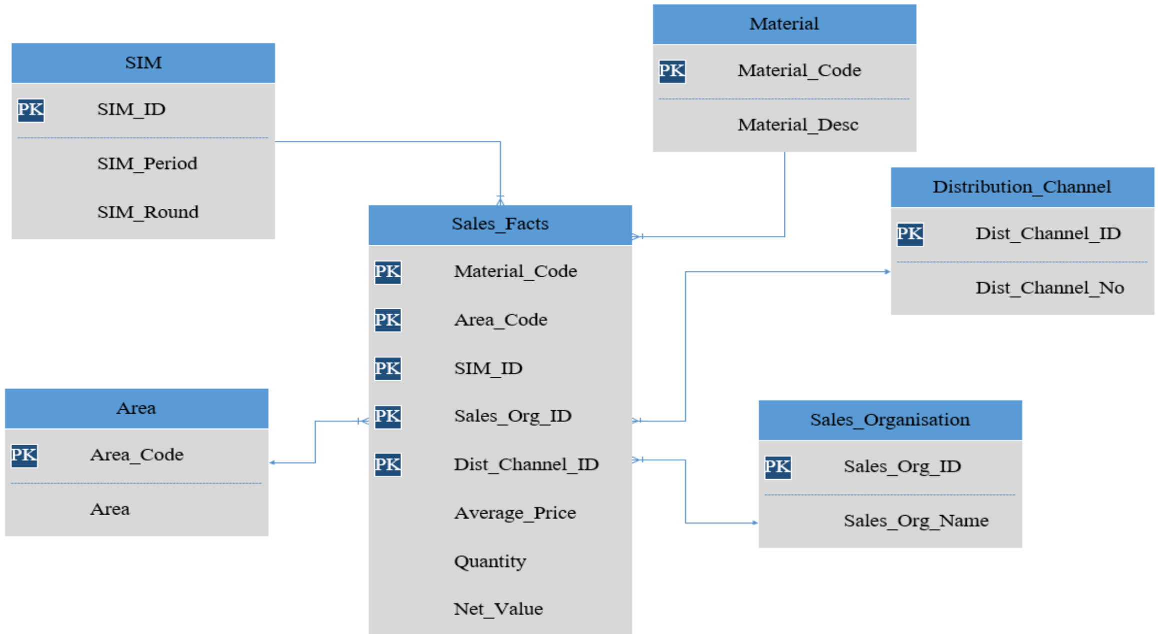

Figure 10. Sales in Multi-dimensional Data Modeling of SAP HANA Data Warehouse

Students model and pivot the multi-dimensional data in the SAP HANA analytical views in this module. This helps students build a better understanding of enterprise data structures. Module 4 is designed for students to achieve learning objectives on the SAP HANA data warehousing: 1) modeling and pivoting ERP multi-dimensional data, 2) connecting to analytical tools, and 3) creating SQL for data warehousing. Students write and execute SQL statements in SAP HANA through OData Services access (see the example in Module 3).

Module 5. Developing the analytics project on data generated from ERPsim Games

In this phase, students developed an analytics project report in teams. In so doing, students download enterprise datasets from the ERPsim. To help students work on the project, the instructor provides the learning resources and materials as project guidelines. The guidance on the resources and materials is demonstrated and discussed in class and during virtual or face-face office hours. The instructor also works together with student teams through the project journey.

Student Learning Assessment

To verify the effectiveness of the developed framework, the author created an SAP HANA learning survey instrument based on SAP HANA learning objectives at a metropolitan public university. The survey includes assessing student HANA knowledge, skills, and learning experiences. The survey was conducted in the database and data warehousing classes in 2019-2022. Ninety-two students responded to the study. Eighty-six responses were complete and valid. The questionnaire survey measures students perceived SAP HANA knowledge, skills, and learning experiences using a Likert scale from 1 to 5. The author performed descriptive analysis on the collected valid data. Figure 11 shows the result of the survey.

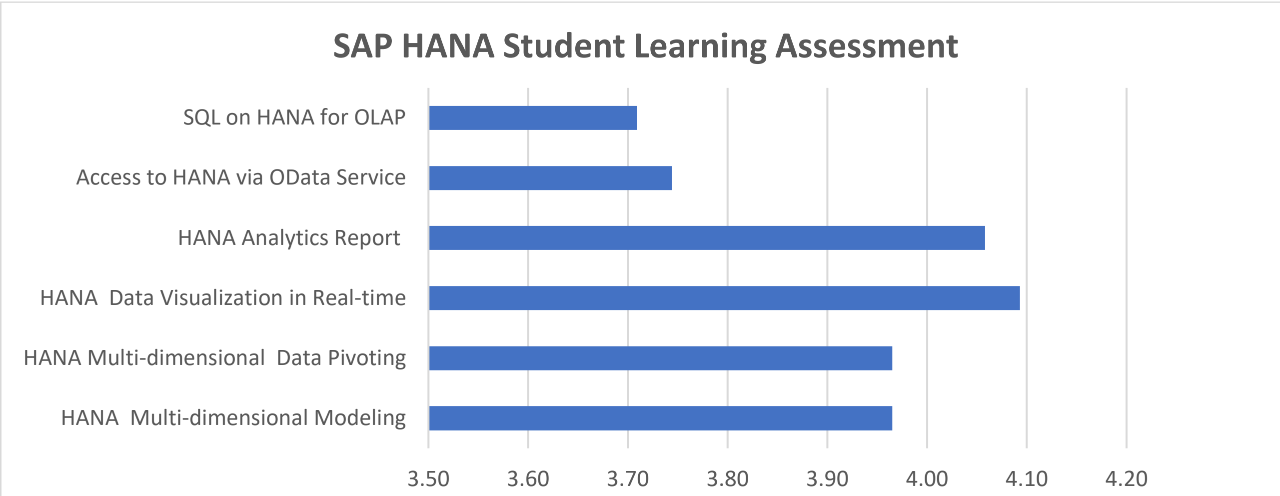

Figure 11 SAP HANA Student Learning Assessment (86 responses)

Students are agreeable that they have learned SPA HANA. The findings of the six questions in the survey ranged from 3.7 to 4.10. Question 3 and Question 4 ratings in the survey are higher because students participated in SAP ERPsim game implementation and enjoyed data visualization on HANA. They developed HANA Analytics Reports on the data they generated from the ERPsim. It is not surprising that Question 1 and Question 2 ratings are a little lower because students need more time to practice SQL queries on HANA, and they only use OData Service as a middleware black box and don't see the algorithm inside. The students have no prerequisite knowledge in SAP HANA before the classes. Thus, no pretest assessment was conducted.

Discussion

The author teaches database and data warehouse Systems courses at graduate and undergraduate levels. Oracle DBMS is utilized in a relational database and object-relational database parts. SAP HANA is used in teaching the data warehouse part in the second part of the courses. The data warehouse concepts are introduced in the textbook without specifying any data warehousing systems. The author utilizes SAP HANA in ERPsim as a tool for students to have hands-on learning experience on data warehousing. SAP HANA learning materials are downloaded from SAP University Alliances Learning Portal as supplemental learning materials. The assessment results demonstrate that students achieved positive learning outcomes in integrating SAP HANA into data warehousing learning.

By adopting the teaching data warehousing with the SAP HANA framework, the IS educators can teach the fundamental data warehousing concepts and skills and real-time data visualization and analytics techniques in the course. IT trend shifts from developing traditional functional applications to enterprise systems. In-memory column store data platform has become the trend in the IT industry. The significant impacts of the framework will be on students' recruitment, job placement, and career development.

Paper Contribution

The paper's contribution is the integration of SAP HANA into teaching data warehousing. The author successfully mapped Three Data Warehouse Architectures (Hoffer, Ramesh, & Topi, 2022) to SAP data warehouse systems in different versions and identified the different data warehouse architectures in traditional SAP BW and SAP HANA on Cloud.

The study provides an important and practical guideline to teach data warehousing with hands-on experiences. This study offers an innovative interdisciplinary approach to teaching database and data warehousing courses and meeting the global shortage of next-generation business leaders and IT professionals in the digital business world.

Implementation Recommendations

For illustration purposes, the study currently is SAP HANA orientated. Before the class, the author recommends information systems instructors test the connection of multiple systems: SAP Server, ERPsim Server, and HANA on cloud and data analytical tools. The author also recommends the IS instructors closely follow the update on the SAP Learning Portal and ERPsim Learning Portal to keep themselves updated with new technology and learning materials.

For non-SAP schools, instructors can match the data warehouse product they use to teach (i.e., Postgres, Oracle or AWS) with the Three Data Warehouse Architectures (Hoffer, Ramesh, & Topi, 2022). Students can have a better understanding if they have the opportunity to have hands-on experience with the specific data warehouse software.

Institutions can choose other data warehouses software packages such as Amazon, Oracle, or Microsoft. Thus, the general usability of the study is evident. Learning one data warehousing software package helps students to understand the fundamentals of other data warehousing software packages. Meanwhile, data warehousing vendors always invent new technology and analytical tools, and information systems educators need to keep learning and bring them into the classroom.

Feature Study

The author will conduct a comparative study on SAP HANA and the newly released version of Oracle DBMS and investigate the revolutionary and evolutionary changes in data modeling, data stores, and analytical performance in these two products. Meanwhile, the proposer will also develop guidelines for IS educators to teach this dynamically changing technology. The significant findings of such research will benefit academic researchers, industry practitioners, and IS educators.

References

Hoffer, J., Ramesh, V, and Topi, H. (2022), Modern Database Management, 13th Edition, Pearson Publisher. Website: www.pearsonhighered.com/hoffer

Inmon, B. (1997). Iterative Development in the Data Warehouse, *DM Review*, 7(11), 16-17.

Inmon, W. (2000). The Problem with Dimensional Modeling, *DM Review*, 10(5), 68–70.

Léger et al. (2017). ERPsim Participant's Guide, ERPsim Lab, HEC Montreal, Canada, https://erpsim.hec.ca (Login user id and password are needed to access the guide)

Marco, D. (2000) Building and Managing the Meta Data Repository: A Full Life-Cycle Guide. New York: Wiley.

Meyer, A. (1997). The Case for Dependent Data Marts. *DM Review,* 7(7), 17–24.

Michon, et al. (2017). ERPsim for SAP HANA Reference Guide, ERPsim Lab, HEC Montreal, Canada.

SAP University Alliances Learning Portal (2022): https://learninghub.sap.com (Login user id and password are needed to access the guide)